

Research Article

Estimation of Daily Reference Evapotranspiration Using Random Forest Optimized by Genetic Algorithm

S Monavar Sabegh¹, D Zare Haghi^{*2}, S Samadianfard³, MR Neishabouri⁴, F Mikaeili⁵

Received: November 2, 2021

Accepted: April 4, 2022

Revised: February 16, 2022

Published online: December 22, 2023

1- Ph.D. Student, Dept. of Soil Sci. Eng., Faculty of Agric., University of Tabriz, Iran

2- Assoc. Prof., Dept. of Soil Sci. Eng., Faculty of Agric., University of Tabriz, Iran

3- Assoc. Prof., Dept. of Water Sci. Eng., Faculty of Agric., University of Tabriz, Iran

4- Prof., Dept. of Soil Sci. Eng., Faculty of Agric., University of Tabriz, Iran

5- M.Sc. student, Dept. of Water Sci. Eng., Faculty of Agric., University of Tabriz, Iran

*Corresponding Author, Email: davoudhaghi@tabrizu.ac.ir

Abstract

Background and Objectives

Reference evapotranspiration (ET_0) is an important parameter in the interactions among soil, vegetation, atmosphere, surface energy and water. Direct measurement of evapotranspiration values is costly and time consuming. On the other hand, modeling this complex process in which many variables interact with each other is not feasible without considering multiple assumptions. In this regard, the FAO Penman-Monteith method is used in a wide range of climatic and environmental conditions. One of the weaknesses of FAO Penman-Monteith method is its dependence on various meteorological parameters. Therefore, it is necessary to implement methods with lower meteorological variables that can estimate ET_0 with suitable accuracy. Thus, in the present study, an attempt was made to estimate ET_0 with acceptable accuracy using machine learning models.

Methodology

In the present study, daily meteorological parameters in the time period of 2000-2020 including maximum and minimum air temperature (T_{max} , T_{min}), mean temperature (T), wind speed (U_2), average relative humidity (RH), maximum and minimum relative humidity (RH_{max} , RH_{min}) and sunshine hours (n) were obtained on a daily basis in three stations of East Azerbaijan province (Tabriz, Sarab, and Maragheh). Moreover, six scenarios were defined as input combinations. Then, using random forest (RF) method in two cases: Single random forest and using the genetic algorithm (GA) to optimize its effective parameters with considering the FAO Penman-Monteith model as a basis, the machine learning models were calibrated and validated for estimating ET_0 values at studied stations. Furthermore, the performance of empirical equations in three groups based on temperature (Hargreaves, Blaney-Criddle and Romanenko), radiation (Irmak) and mass transfer (Meyer) were also investigated. It should be noted that 75% of the data were considered for calibration and 25% for the validation of machine learning methods. Finally, using the statistical criteria of correlation coefficient (CC), scattered index (SI) and Willmott's Index of agreement (WI), a suitable machine learning method was introduced to estimate the reference evapotranspiration. Also, the most suitable combination of meteorological parameters for ET_0 estimation was suggested.

Findings

The obtained results showed that in all studied stations, the scenario 6 has the best performance, either in the case of single random forest (RF) or in the case of random forest optimized by genetic algorithm (GA-RF). Meteorological parameters of this scenario include minimum and maximum air temperature, minimum and



maximum relative humidity, sunshine hours and wind speed. By optimizing the RF-6 parameters with the genetic algorithm at Tabriz Station, the statistical criteria were improved (CC from 0.990 to 0.991, SI from 0.103 to 0.098). At Sarab Station, the CC was increased from 0.980 to 0.982, the SI was decreased from 0.140 to 0.132 and the WI was increased from 0.989 to 0.990. At Maragheh Station, CC was increased from 0.990 to 0.991, SI was decreased from 0.103 to 0.098 and WI remained unchanged at 0.995. In general, the decreasing trend of the scattered index for RF method from scenarios 1 to 6 can be understood by increasing the input parameters of the random forest method. Among the three groups of empirical methods based on air temperature, radiation and mass transfer for estimating ET_0 , the best performance was seen for the

Blaney-Criddle method based on air temperature. In all studied stations, the GA-RF model showed better performance than the empirical methods. Also, GA-RF-5 with similar meteorological parameters with Blaney-Criddle method provided accurate ET_0 estimations.

Conclusion

Determining the amount of daily evapotranspiration and consequently accurate estimation of water requirement of plants provide the basis for proper designing of irrigation systems by reducing installation costs and providing a suitable program for the use of water resources in the agriculture sector. Therefore, in the present study, meteorological data from Tabriz, Sarab and Maragheh Stations were used to evaluate the ability of machine learning methods including RF and GA-RF to estimate the values of reference evapotranspiration. The results showed the high accuracies of RF-6 and GA-RF-6 for all studied stations and Blaney-criddle among the empirical models. In a more detailed look, the genetic algorithm had positive effects on increasing the model accuracies by reducing scattered index of GA-RF scenarios 1, 4, 5 and 6 in Tabriz and Maragheh Stations as well as scenarios 1, 5 and 6 at Sarab Station. Finally, it can be concluded that both RF and GA-RF models provided the most accurate estimates of daily reference evapotranspiration in the East Azerbaijan province.

Keywords: East Azerbaijan, Genetic algorithm, Optimized, Random Forest, Reference evapotranspiration

مقاله پژوهشی

برآورد تبخیر- تعرق مرجع روزانه با استفاده از روش جنگل تصادفی بهینه شده با الگوریتم ژنتیک

ساناز منور سابق^۱، داوود زارع حق^{۲*}، سعید صمدیان فرد^۳، محمدرضا نیشابوری^۴، فاطمه میکائیلی^۵

تاریخ دریافت: ۱۴۰۰/۰۸/۱۱

تاریخ پذیرش: ۱۴۰۱/۰۱/۱۵

تاریخ ویرایش: ۱۴۰۰/۱۱/۲۷

تاریخ انتشار آنلاین: ۱۴۰۲/۱۰/۰۱

- ۱- دانشجوی دکتری، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تبریز
 - ۲- دانشیار، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تبریز
 - ۳- دانشیار، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز
 - ۴- استاد، گروه علوم و مهندسی خاک، دانشکده کشاورزی، دانشگاه تبریز
 - ۵- دانشجوی کارشناسی ارشد، گروه علوم و مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز
- * مسئول مکاتبات، پست الکترونیکی: davoudhaghi@tabrizu.ac.ir

چکیده

تبخیر- تعرق یک متغیر مهم در فعل و انفعالات بین خاک، پوشش گیاهی، جو، انرژی سطح زمین و آب است. از طرفی، اندازه‌گیری آن از طریق روش‌های مستقیم، هزینه و زمان زیادی می‌طلبد. هدف از پژوهش حاضر، بررسی توانایی روش جنگل تصادفی (RF) در دو حالت منفرد و بهینه شده با الگوریتم ژنتیک (GA-RF) می‌باشد. بدین منظور، داده‌های روزانه برخی از متغیرهای هواشناسی اثرگذار بر پدیده تبخیر- تعرق در دوره آماری ۲۰ ساله (۱۴۰۰-۱۳۸۰) در سه ایستگاه تبریز، سراب و مراغه واقع در استان آذربایجان شرقی جمع‌آوری شد. سپس شش سناریو ترکیبی از متغیرهای هواشناسی برای واسنجی و صحت‌سنجی مدل‌های مذکور مد نظر قرار گرفتند. علاوه بر این، عملکرد سه گروه از روش‌های تجربی برآورد کننده تبخیر- تعرق مرجع نیز مورد بررسی قرار گرفت. در نهایت با استفاده از معیارهای آماری، کارایی روش‌ها مورد ارزیابی قرار گرفت. نتایج نشان داد که به‌منظور تخمین ET_0 با استفاده از متغیرهای هواشناسی کمتر سناریو ۴ با شاخص پراکندگی ۰/۱۳۱ در ایستگاه تبریز، ۰/۱۷۱ در ایستگاه سراب و ۰/۱۳۴ در ایستگاه مراغه دقت بالایی دارد. همچنین سناریو ۲ در ایستگاه‌های تبریز، سراب و مراغه به‌ترتیب با شاخص پراکندگی ۰/۱۸۴، ۰/۲۲۰ و ۰/۱۷۲ با دقت قابل قبولی می‌تواند مورد استفاده قرار گیرد. در حالت مقایسه جنگل تصادفی عملکرد بهتری نسبت به روش‌های تجربی در ایستگاه‌های مورد مطالعه نشان داد. در نهایت، استفاده از روش جنگل تصادفی به‌منظور برآورد دقیقی از میزان تبخیر- تعرق مرجع در استان آذربایجان شرقی پیشنهاد گردید.

واژه‌های کلیدی: آذربایجان شرقی، الگوریتم ژنتیک، بهینه شده، تبخیر- تعرق مرجع، جنگل تصادفی

مقدمه

فرآیند تبخیر- تعرق (ET)^۱ توسط عوامل متعددی که در انتقال آب از خاک به گیاه و نهایتاً به اتمسفر و تبخیر از منطقه تحت کشت مؤثرند، کنترل می‌شود. مدل‌سازی چنین فرآیند پیچیده‌ای که بسیاری از پارامترهای مؤثر بر آن بر یکدیگر نیز اثر متقابل می‌گذارند، امری بسیار مشکل بوده که بدون فرض‌های متعدد جهت ساده نمودن موضوع امکان‌پذیر نمی‌باشد. در مدل‌های برآورد ET همواره میزان پیچیدگی مدل و حجم داده‌های مورد نیاز آن مورد توجه بوده است. پژوهشگران سعی نموده‌اند جهت ساده‌سازی، مدل‌هایی ارائه نمایند که بتوان بر اساس داده‌های هواشناسی، تبخیر- تعرق مرجع (ET₀)^۲ را برآورد کرد. یکی از اصلی‌ترین فوائد معرفی ET₀ حذف عوامل متعدد مرتبط به گیاه و خاک بوده که بر فرآیند تبخیر- تعرق اثر می‌گذارند. تعیین میزان نیاز آبی گیاه می‌تواند به طریق مستقیم و با استفاده از تکنیک‌های هواشناسی که مبتنی بر تعادل انرژی و نیز انتقال جریان جرم بخار می‌باشند (روش آیرودینامیک، نسبت باون^۳)، صورت پذیرد. به‌کارگیری لایسیمتر امکان اندازه‌گیری مقدار تبخیر- تعرق را براساس اندازه‌گیری برخی از مؤلفه‌های تعادل آبی در یک سطح گیاهی کنترل شده فراهم می‌آورد. همچنین تبخیر- تعرق به‌صورت غیرمستقیم نیز تعیین می‌شود، ازجمله گزینه‌های اقتصادی به‌منظور صرفه‌جویی در هزینه‌ها و کاهش زمان مورد نیاز، استفاده از مدل‌های ریاضی مشتمل بر مدل‌های تجربی و فیزیکی می‌باشند که مدل‌های تجربی بر مبنای توابع آماری تقریب بین مقادیر تبخیر- تعرق و متغیرهای هواشناسی استوار هستند. مدل‌های فیزیکی نیز اصول فیزیکی موجود و مربوط به مقدار انرژی لازم برای گرمای نهان تبخیر، جریان بخار آب در خارج از سطح

تبخیر و تأمین بخار آب در سطح تبخیر را که سه عامل مهم تأثیرگذار در تبخیر- تعرق هستند، به‌کار می‌گیرند (کریمی و همکاران ۲۰۱۳). محدودیت منابع آب و خاک و لزوم مدیریت درست این منابع در شرایط گوناگون محیطی، موجب تلاش پژوهشگران برای تعیین ET از راه اندازه‌گیری‌های میدانی و یا مدل‌سازی بوده است (زارع ابیانه و همکاران ۲۰۱۴). تخمین بیش از حد نیاز آبی گیاه ضمن هدر دادن آب آبیاری باعث ماندابی شدن اراضی، شست‌وشوی مواد غذایی خاک و آلوده نمودن منابع آب زیرزمینی می‌شود. از طرفی تخمین کمتر نیز باعث اعمال تنش رطوبتی به گیاه شده و در نتیجه کاهش محصول را به‌همراه خواهد داشت (کوچک زاده و بهمنی ۲۰۰۵).

مدل فائو- پنمن- مانتیث (FPM)^۴ که مدلی ترکیبی است، توسط سازمان خواربار و کشاورزی ملل متحد (FAO)^۵ به‌عنوان روش مرجع استاندارد برای تخمین ET₀ و اعتبارسنجی سایر روش‌ها توصیه می‌شود (آلن و همکاران ۱۹۹۸). روش FPM را می‌توان در طیف گسترده‌ای از شرایط آب‌وهوایی و محیطی، بدون نیاز به هیچ‌گونه واسنجی استفاده نمود. این روش با استفاده از داده‌های لایسیمتری در طیف وسیعی از شرایط آب و هوایی مورد تأیید قرار گرفته است (لاندراس و همکاران ۲۰۰۸). وابستگی به متغیرهای متعدد هواشناسی ضعف این روش محسوب می‌شود. بنابراین استفاده از این معادله در همه زمان‌ها و مکان‌ها میسر نمی‌باشد. لذا استفاده از روش‌هایی که با تعداد کمی از متغیرهای هواشناسی بتواند تبخیر- تعرق مرجع را با دقت بالایی محاسبه نماید ضروری است. بنابراین برای بدست آوردن مقادیر ET₀ از سایر مدل‌های ساده، همراه با مدل‌های محاسبات نرم‌افزاری استفاده می‌شود (شیری و همکاران ۲۰۲۰). مدل‌های تجربی دارای ضرایبی هستند که معرف شرایط

4. FAO Penman-Monteith

5. Food and Agriculture Organization of the United Nations

1. Evapotranspiration

2. Reference Evapotranspiration

3. Bowen Ratio

منطقه‌ای است که مدل در آن واسنجی شده لذا دو مشکل عمده دارند: نخست عمومیت ندارند و به بیان دیگر در شرایط خاص توسعه یافته‌اند. دوم اینکه برخی از این مدل‌ها به اطلاعات زیادی نیاز دارند (پیری و پوزن ۲۰۱۷).

روش‌های هوش مصنوعی (AI)^۱ بهتر از روش‌های کلاسیک قادر به تخمین رفتار سیستم در همه مراحل عملکرد می‌باشند و نیز دارای دقت و سرعت بهتری در تخمین ET_0 در مقایسه با روش‌های کلاسیک می‌باشند (بیات ورکشی و همکاران، ۲۰۰۹). به‌وسیله داده‌کاوی^۲، داده‌های موجود مورد تحلیل قرار می‌گیرند تا روندهای احتمالی، ارتباطات غیرمحسوس و الگوهای مخفی از بین انبوه داده‌ها شناسایی شوند. داده‌کاوی از الگوریتم‌های مختلفی مانند روش جنگل تصادفی (RF)^۳، درخت‌های تصمیم^۴، رگرسیون بردار پشتیبان (SVR)^۵، شبکه‌های عصبی مصنوعی (ANN)^۶ و غیره بهره می‌گیرد. با در نظر گرفتن مزایای الگوریتم‌های داده‌کاوی، ازجمله امکان وارد کردن تعداد بسیار زیادی متغیر به مدل، عدم نیاز به فرضیاتی مانند نرمال بودن و توانایی یافتن پارامترها حتی باوجود داده‌های گمشده، می‌توان این روش‌ها را ابزاری مناسب برای تحلیل داده‌هایی با حجم بالا در نظر گرفت (هاشمیان و همکاران ۲۰۱۷). فنگ و همکاران (۲۰۱۷) دو مدل داده‌محور جنگل تصادفی و شبکه‌های عصبی رگرسیون تعمیم‌یافته (GRNN)^۷ را برای تخمین ET_0 در دو ایستگاه واقع در جنوب غربی چین ارزیابی کردند. از دو ترکیب ورودی، شامل داده‌های کامل و فقط داده‌های دما و تابش فرازمینی (R_a) و آزمون k-fold برای آزمایش عملکرد مدل‌ها با توجه به معیارهای زمانی و مکانی

استفاده کردند و عملکرد کلی بهتر RF را گزارش دادند. در مطالعه عَنابی میلانی و نیشابوری (۲۰۱۸) تبخیر-تعرق مرجع با روش بیلان آبی خاک با استفاده از لایسیمتر برای دوره‌های ده روزه و ماهانه در دشت تبریز تعیین گردید و با نتایج به‌دست آمده از برخی روابط تجربی مقایسه شد و با توجه به نتایج در کل بهترین روش برای برآورد تبخیر-تعرق مرجع در دشت تبریز روش FPM شناخته شد. نتایج تحقیق فن و همکاران (۲۰۱۸) در هشت ایستگاه هواشناسی چین نشان داد که تابش خورشیدی نسبت به رطوبت نسبی و سرعت باد برای برآورد ET_0 روزانه در مناطق موسمی گرمسیری و نیمه گرمسیری مهم‌تر است. همچنین مدل‌های یادگیری ماشین با استفاده از حداقل و حداکثر دمای هوای روزانه، رطوبت نسبی، سرعت باد و تابش خورشیدی، تخمین ET_0 روزانه رضایت‌بخشی را در مناطق معتدل قاره، فلات کوهستان و مناطق موسمی معتدل چین به‌دست آورده‌اند. وانگ و همکاران (۲۰۱۹) در مطالعه‌ای با هدف بررسی توانایی تعمیم الگوریتم جنگل تصادفی در مدل‌سازی ET_0 با ترکیبات ورودی مختلف و مقایسه این الگوریتم با روش برنامه‌ریزی بیان ژن^۸ با استفاده از داده‌های بیست و چهار ایستگاه هواشناسی در جنوب غربی چین نتیجه گرفتند که RF در مدل‌سازی ET_0 با متغیرهای کامل و ناقص هواشناسی با موفقیت عمل می‌کند و به‌طور کلی، هرچه متغیرهای بیشتری توسط یک مدل استفاده شوند، شانس بیشتری برای عملکرد بهتر دارد. دا سیلوا جونیور (۲۰۱۹) الگوریتم‌های درونیابی مکانی داده‌های تبخیر-تعرق را از نظر دقت و عملکرد ارزیابی نمود و دریافت که روش جنگل تصادفی نسبت به روش‌های مرسوم درونیابی همچون روش معکوس وزنی فاصله و کریجینگ معمولی نتایج بهتری ارائه می‌دهد. سیاسر و همکاران (۲۰۲۰) به بررسی توانایی مدل‌های خطی تعمیم‌یافته، جنگل تصادفی و درخت گرادیان تقویتی در

1. Artificial Intelligence
2. Data Mining
3. Random Forest
4. Decision Trees
5. Support Vector Regression
6. Artificial Neural Networks
7. Generalized Regression Neural Networks

آذربایجان شرقی است که دقت آن‌ها با روش فائو-پنمن-مانتیت ارزیابی می‌گردد. در نهایت با بررسی و ارزیابی آماری هر یک از روش‌های مذکور با روش‌های تجربی بر پایه دمای هوا شامل هارگریوز-سامانی^۳، بلانی-کریدل^۴ و روماننکو^۵، بر پایه تابش ایرماک^۶ و بر پایه انتقال جرم مییر^۷ دقیق‌ترین روش پیشنهاد می‌شود.

مواد و روش‌ها

منطقه مورد مطالعه

استان آذربایجان شرقی دارای آب و هوای سرد کوهستانی بوده و مساحت آن ۴۵۴۹۱ کیلومتر مربع است. موقعیت مکانی ایستگاه‌های مورد مطالعه در شکل ۱ و مشخصات جغرافیایی آن‌ها در جدول ۱ قابل مشاهده می‌باشد. در یک دوره آماری ۲۰ ساله (۱۴۰۰-۱۳۸۰) داده‌های هواشناسی شامل دمای هوای حداکثر و حداقل (T_{max}, T_{min})، دمای میانگین (T)، سرعت باد در ارتفاع دو متری (U_2)، رطوبت نسبی میانگین (RH)، رطوبت نسبی حداکثر و حداقل (RH_{max}, RH_{min}) و ساعات آفتابی (n) در مقیاس روزانه از اداره کل هواشناسی استان آذربایجان شرقی اخذ شد. مشخصات آماری متغیرهای هواشناسی مورد استفاده برای ایستگاه‌های تبریز، سراب و مراغه در جدول ۲ ارائه شده است.

شبیه‌سازی تبخیر-تعرق مرجع در دشت سیستان پرداختند که بهترین عملکرد مربوط به درخت گرادیان تقویتی بود. در این مطالعه برای مدل جنگل تصادفی بهترین الگو شامل ترکیب متغیرهای دمای حداکثر، رطوبت میانگین، سرعت باد و تبخیر از تشت بود. همچنین دمای هوا مهم‌ترین متغیر مؤثر در الگوهای انتخابی برای هر سه مدل شناخته شد. سلام و همکاران (۲۰۲۰) نتیجه گرفتند که مدل RF برای پیش‌بینی تبخیر-تعرق مرجع روزانه با استفاده از داده‌های هواشناسی محدود در منطقه مرطوب بنگلادش به دلیل دقت بالا، پایداری بیشتر و هزینه محاسباتی کمتر، بهتر از سایر مدل‌های برآوردکننده ET_0 روزانه عمل می‌کند. تجزیه و تحلیل RF نشان داد که سرعت باد و تابش خورشیدی تأثیرگذارترین متغیرها در برآورد ET_0 می‌باشند. صمدیان فرد و همکاران (۲۰۲۰) توانایی روش‌های داده‌محور رگرسیون فرآیند گاوسی (GPR)^۱، رگرسیون بردار پشتیبان و جنگل تصادفی در تخمین تبخیر-تعرق مرجع را در نه ایستگاه شمالی کشور مورد بررسی قرار دادند و مدل‌های GPR و SVR توانایی بالاتری در تخمین هر چه دقیق‌تر تبخیر-تعرق در مناطق مرطوب داشتند. در تحقیق چن و همکاران (۲۰۲۰)، نتایج نشان داد که هنگام استفاده از روش‌های RF و SVR برای پیش‌بینی تبخیر-تعرق مرجع روزانه، داده‌های هواشناسی تابش خورشیدی و دمای هوا اساسی و بسیار مهم هستند. همچنین با مقایسه بین دو روش RF زمانی توصیه می‌شود که داده‌های ورودی نسبتاً کامل باشند یا فقط فاقد داده‌های خاک و گیاهان باشند.

هدف از تحقیق حاضر، برآورد تبخیر-تعرق مرجع با استفاده از روش جنگل تصادفی به عنوان یکی از روش‌های داده‌محور در دو حالت منفرد و جنگل تصادفی بهینه شده با الگوریتم ژنتیک (GA-RF)^۲ در سه ایستگاه تبریز، سراب و مراغه واقع در استان

3. Hargreaves-Samani

4. Blaney-Criddel

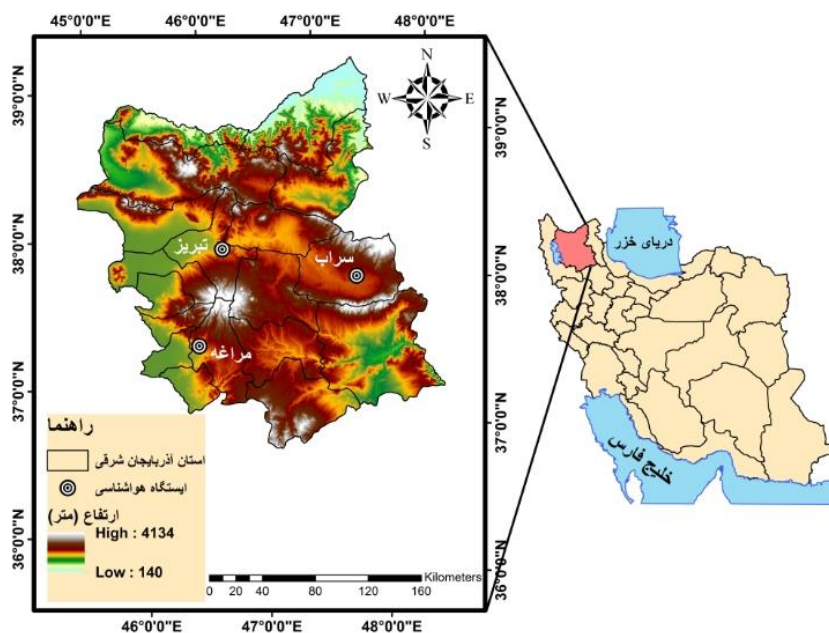
5. Romanenko

6. Irmak- R_n

7. Meyer

1. Gaussian Process Regression

2. Random Forest Optimized by Genetic Algorithm



شکل ۱- پراکنش ایستگاه‌های مورد مطالعه در آذربایجان شرقی، ایران.

جدول ۱- مشخصات جغرافیایی ایستگاه‌های سینوپتیک مورد مطالعه.

ایستگاه	طول جغرافیایی (E)	عرض جغرافیایی (N)	ارتفاع (m)
تبریز	۴۶° ۱۷'	۳۸° ۵'	۱۳۵۱
سراب	۴۷° ۳۲'	۳۷° ۵۶'	۱۶۵۰
مراغه	۴۶° ۱۶'	۳۷° ۲۴'	۱۴۷۷

جدول ۲- مشخصات آماری داده‌های مورد استفاده در ایستگاه‌های مورد مطالعه.

ایستگاه	متغیر	میانگین	حداقل	حداکثر	انحراف معیار	ضریب تغییرات	چولگی
تبریز	T_{min} (°C)	۸/۰۹۸	-۱۸/۰۰۰	۳۰/۰۰۰	۹/۴۸۸	۱/۱۷۲	-۰/۱۴۲
	T_{max} (°C)	۱۹/۷۰۲	-۷/۹۰۰	۴۱/۰۰۰	۱۱/۲۶۶	۰/۵۷۲	-۰/۱۷۰
	n (hr)	۷/۹۶۶	۰/۰۰۰	۱۳/۹۰۰	۳/۷۶۹	۰/۴۷۳	-۰/۷۲۹
	U_2 (m s ⁻¹)	۳/۵۰۷	۰/۰۰۰	۱۱/۰۰۰	۱/۵۶۴	۰/۴۴۶	۰/۸۳۲
	RH _{min} (%)	۳۲/۳۶۱	۲/۰۰۰	۹۲/۰۰۰	۱۷/۱۹۸	۰/۵۳۱	۰/۸۲۳
	RH _{max} (%)	۶۷/۷۳۲	۱۵/۰۰۰	۱۰۰/۰۰	۱۸/۵۲۵	۰/۲۷۳	-۰/۳۱۹
	ET ₀ (mm day ⁻¹)	۴/۲۴۲	۰/۰۶۱	۱۴/۳۷۲	۲/۹۹۰	۰/۷۰۵	۰/۴۸۸
سراب	T_{min} (°C)	۲/۳۰۵	-۳۰/۰۰۰	۲۳/۴۰۰	۸/۶۲۷	۳/۷۴۲	-۰/۵۷۶
	T_{max} (°C)	۱۶/۹۴۱	-۱۰/۰۰۰	۳۷/۰۰۰	۱۰/۹۳۷	۰/۶۴۵	-۰/۲۱۰
	n (hr)	۸/۱۲۲	۰/۰۰۰	۱۳/۹۰۰	۳/۷۲۶	۰/۴۵۹	-۰/۶۵۴

۱/۳۳۳	۰/۶۲۵	۱/۸۶۶	۱۵/۶۲۵	۰/۰۰۰	۲/۹۸۵	U_2 (m s ⁻¹)
۰/۵۸۸	۰/۵۲۶	۱۹/۱۹۳	۹۷/۰۰۰	۴/۰۰۰	۳۶/۵۱۶	RH _{min} (%)
-۱/۰۴۰	۰/۱۴۸	۱۲/۲۵۱	۱۰۱/۰۰۰	۳۰/۰۰۰	۸۳/۰۳۴	RH _{max} (%)
۰/۳۶۰	۰/۶۸۱	۲/۲۱۹	۹/۳۰۳	۰/۱۱۳	۳/۲۵۷	ET ₀ (mm day ⁻¹)
-۰/۰۷۳	۱/۰۴۳	۹/۱۶۸	۲۸/۶۰۰	-۱۶/۴۰۰	۸/۷۸۷	T _{min} (°C)
-۰/۱۰۶	۰/۵۳۷	۱۱/۱۶۸	۴۲/۰۰۰	-۴/۶۰۰	۲۰/۸۰۹	T _{max} (°C)
-۰/۷۰۶	۰/۴۸۹	۳/۹۹۱	۱۳/۶۰۰	۰/۰۰۰	۸/۱۵۸	n (hr)
۰/۹۰۶	۰/۴۳۲	۱/۵۶۲	۱۰/۷۵۰	۰/۰۰۰	۳/۶۱۵	U_2 (m s ⁻¹) مراغه
۰/۹۱۲	۰/۵۶۳	۱۸/۸۶۵	۹۴/۰۰۰	۲/۰۰۰	۳۳/۵۲۵	RH _{min} (%)
-۰/۱۷۲	۰/۳۲۹	۲۱/۵۰۰	۱۰۱/۰۰۰	۱۸/۰۰۰	۶۵/۳۸۴	RH _{max} (%)
۰/۴۸۶	۰/۷۱۴	۳/۱۸۷	۱۳/۸۹۰	۰/۱۴۱	۴/۴۶۶	ET ₀ (mm day ⁻¹)

روش هارگریوز- سامانی یکی از ساده‌ترین رابطه‌ها برای برآورد تبخیر- تعرق مرجع است و از دقت نسبتاً خوبی در مناطق خشک و نیمه‌خشک برخوردار می‌باشد و در آن به دمای هوای حداقل و حداکثر و تابش فرازمینی نیاز می‌باشد (صمدیان فرد و همکاران ۲۰۲۰). روش بلانی- کریدل یکی از قدیمی‌ترین روش‌های تخمین تبخیر- تعرق مرجع است که در آن a و b ضرایب اقلیمی و P ضریب مربوط به طول روز یا درصد سالانه تابش آفتاب در ماه که به‌صورت روزانه توصیف شده است. روش مییر از مفهوم انتقال گردابی بخار آب از سطح تبخیر به اتمسفر استفاده کرده و بر اساس قانون گاز دالتون هستند.

جدول ۳- معادلات و متغیرهای ورودی آن‌ها.

مدل	متغیرهای ورودی	فرمول
فائو- پنمن- مانتیت (آن و همکاران، ۱۹۹۸)	$G, \gamma, T, U_2, e_a, e_d, \Delta R_n$	$ET_0 = \frac{0.408\Delta(R_n - G) + \gamma \frac{890}{T + 273} U_2 (e_a - e_d)}{\Delta + \gamma(1 + 0.34U_2)}$
هارگریوز- سامانی (۱۹۸۵)	R_a, T, T_{max}, T_{min}	$ET_0 = 0.0023R_a (T + 17.8)\sqrt{T_{max} - T_{min}}$

در این پژوهش، روش استاندارد FPM برای محاسبه تبخیر- تعرق مرجع به‌عنوان روش مبنا مورد استفاده قرار گرفت و ۷۵ درصد داده‌ها برای واسنجی و ۲۵ درصد برای صحت‌سنجی روش‌های یادگیری ماشین مورد مطالعه در نظر گرفته شد. جدول ۳ خلاصه‌ای از معادلات به‌کار گرفته شده به‌همراه متغیرهای هواشناسی مورد نیاز برای محاسبه تبخیر- تعرق مرجع در مقیاس روزانه (مطابق سناریوهای روش‌های یادگیری ماشین) را نشان می‌دهد. همانطور که اشاره شد معادله فائو- پنمن- مانتیت به‌عنوان دقیق‌ترین روش برآورد تبخیر- تعرق مرجع برای اعتبارسنجی سایر روش‌ها انتخاب شد.

$$ET_0 = a + b[P(0.46 T + 8.13)]$$

$$a = 0.0043(RH_{min}) - \frac{n}{N} - 1.41$$

$$b = 0.82 - 0.0041(RH_{min}) + 1.07\left(\frac{n}{N}\right) + 0.066(U_2) - 0.006(RH_{min}) \times \left(\frac{n}{N}\right) - 0.0006(RH_{min})(U_2)$$

بلانی-کریدل
T, RH_{min}, n, N, U₂
(۱۹۵۰)

$$ET_0 = 0.00006 \times (T + 25)^2 \times (100 - RH)$$

رومانکو
T, RH
(۱۹۶۱)

$$ET_0 = 0.489 + 0.289R_n + 0.023T$$

ایرماک-R_n
R_n, T
(۲۰۰۳)

$$ET_0 = (3.75 + 0.5026 \times U_2) \times (e_a - e_d)$$

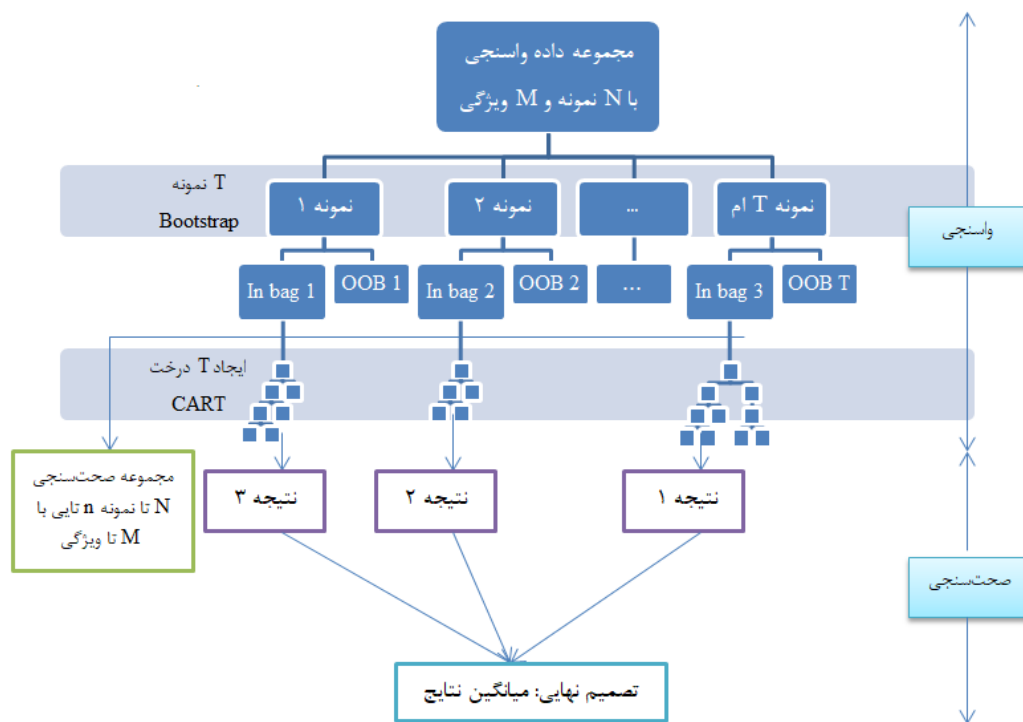
مییبر (۱۹۲۶)
U₂, e_a, e_d

ET_0 ← تبخیر تعریق گیاه مرجع (mm day^{-1})، R_n ← تابش خالص در سطح پوشش گیاهی ($\text{MJ m}^{-2} \text{ day}^{-1}$)، G ← شار گرما به داخل خاک ($\text{MJ m}^{-2} \text{ day}^{-1}$)، γ ← ضریب سایکرومتری ($\text{kPa } ^\circ\text{C}^{-1}$)، T ← متوسط دمای هوا ($^\circ\text{C}$)، U_2 ← سرعت باد در ارتفاع دو متری از سطح زمین (m s^{-1})، e_a ← فشار بخار واقعی (kPa)، e_d ← فشار بخار اشباع (kPa)، Δ ← شیب منحنی فشار بخار ($\text{kPa } ^\circ\text{C}^{-1}$)، R_a ← تابش فرازمینی (mm day^{-1})، T_{max} ← حداکثر دمای هوا ($^\circ\text{C}$)، T_{min} ← حداقل دمای هوا ($^\circ\text{C}$)، RH_{min} ← حداقل رطوبت نسبی (%)، n ← ساعات آفتابی (hr)، N ← تعداد ساعات روشنایی روز (hr)، RH ← متوسط رطوبت نسبی (%).

جنگل تصادفی

در این الگوریتم، از چندین درخت تصمیم استفاده می‌شود که به هرکدام از درخت‌ها، زیرمجموعه‌ای از داده‌ها اختصاص داده می‌شود. این درخت‌ها با داده‌های زیرمجموعه، می‌توانند تصمیم بگیرند و مدل طبقه‌بندی خود را بسازند. تنوع درختان از طریق ایجاد زیرمجموعه‌های مختلف از داده‌های آموزشی، کاهش می‌یابد که اصطلاحاً دسته‌بندی نامیده می‌شود. دسته‌بندی تکنیکی است که از طریق نمونه‌برداری مجدد تصادفی از مجموعه داده‌های اصلی و همراه با جایگزینی انجام می‌شود. در این مرحله هیچکدام از داده‌های انتخاب شده از نمونه‌های ورودی برای تولید زیرمجموعه بعدی، حذف نمی‌شود و بدین

ترتیب واریانس نیز کاهش می‌یابد. از این رو برخی از داده‌ها ممکن است بیش از یک بار در شاخه‌های آموزشی استفاده شود. بنابراین ثبات بیشتری برای مدل به دست می‌آید و مدل را در برابر تغییرات جزئی در داده‌های ورودی قابل اعتمادتر می‌سازد و دقت پیش‌بینی آن را افزایش می‌دهد. هنگامی که یک سری داده جدید برای پیش‌بینی داده می‌شود، هرکدام از الگوریتم‌های یاد گرفته شده قبلی، یک نتیجه را برای پیش‌بینی برمی‌گردانند (شکل ۲). الگوریتم جنگل تصادفی در نهایت می‌تواند با استفاده از رأی‌گیری، آن طبقه‌ای را که بیشترین رأی را آورده است، انتخاب کند و به عنوان طبقه‌ی نهایی برای انجام عملیات طبقه‌بندی قرار دهد (اقنوم و همکاران ۲۰۱۷).

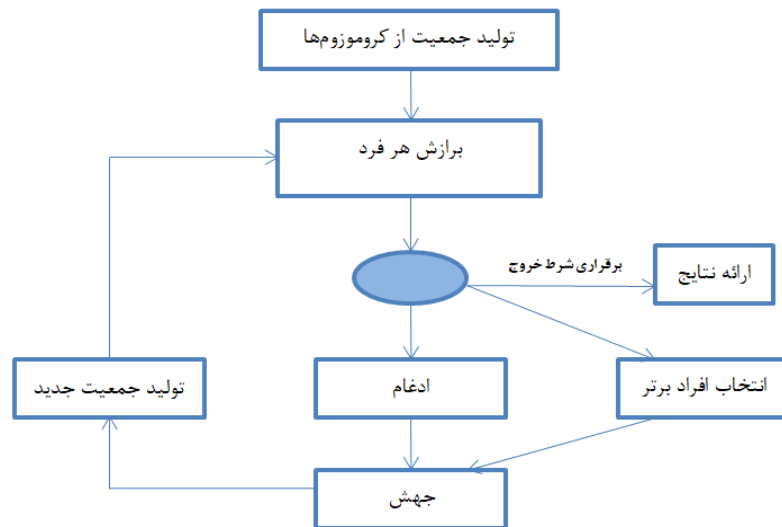


شکل ۲- فلوچارت روش جنگل تصادفی (نوروزی و ندیری ۲۰۱۸).

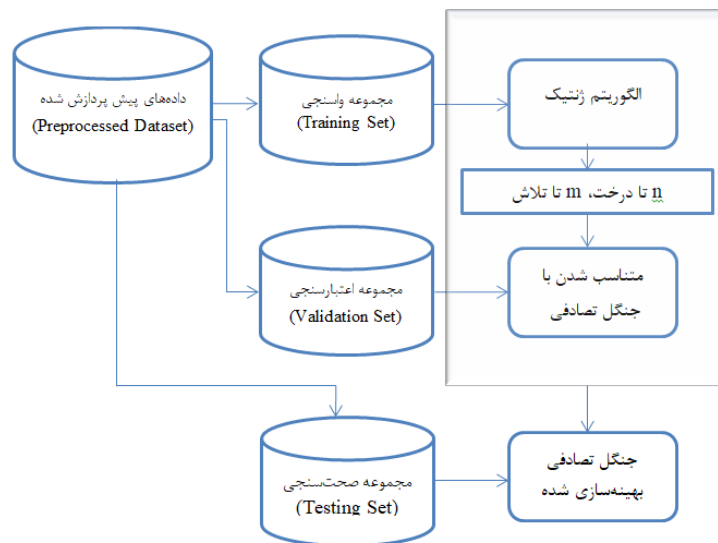
الگوریتم ژنتیک

می‌کنند. هدف از جهش، شروع جست‌وجو در محل دیگری از فضای تصمیم است. از میان کروموزوم‌های قبلی و جدید، کروموزوم‌هایی که شایستگی بهتری دارند، انتخاب می‌شوند و به این ترتیب اندازه جمعیت در نسل بعدی نیز ثابت خواهد ماند (شکل ۳). پس از چندین تکرار، الگوریتم به سمت بهترین کروموزوم همگرا می‌شود که نشان دهنده جواب بهینه یا نزدیک به بهینه خواهد بود (محمدی و همکاران ۲۰۱۹). در پژوهش حاضر جنگل تصادفی با الگوریتم ژنتیک بهینه شده، همانطور که در شکل ۴ نشان داده شده است در ابتدا داده‌های پیش‌پردازش شده به سه قسمت واسنجی، اعتبار سنجی و صحت سنجی تقسیم گردید سپس GA با در نظر گرفتن داده‌های واسنجی، متغیرهای روش RF را با لحاظ نمودن داده‌های اعتبارسنجی متناسب‌سازی نمود و در نهایت روش RF بهینه شده توسط GA با استفاده از داده‌های صحت‌سنجی مورد ارزیابی قرار گرفت.

این الگوریتم زیرمجموعه‌ای از الگوریتم‌های تکاملی و روشی برای یافتن راه‌حل تقریبی برای مسائل بهینه‌سازی است. نخستین گام در پیاده‌سازی الگوریتم ژنتیک، تبدیل جواب مسئله به یک کروموزوم است. تعداد کروموزوم‌های اولیه، اندازه جمعیت نامیده می‌شود. جمعیت اولیه فقط یک بار و در زمان شکل‌گیری اولین نسل الگوریتم ژنتیک به صورت تصادفی ایجاد می‌شود. سپس میزان شایستگی هر یک از کروموزوم‌ها بر اساس تابع هدف محاسبه می‌شود. با استفاده از عملگر انتخاب، تعدادی از کروموزوم‌های جمعیت فعلی به عنوان والدین نسل بعد برگزیده می‌شوند. با تقاطع والدین از طریق عملگر ترکیب، فرزندان نسل آینده که همان راه‌حل‌های جدید هستند، تولید می‌شوند. هدف از این تولید، امکان به وجود آمدن فرزندان بهتر است. سپس با استفاده از عملگر جهش، برخی از ژن‌ها به صورت تصادفی انتخاب و تغییر



شکل ۳- فلوجارت الگوریتم ژنتیک (عمادی و همکاران ۲۰۱۱).



شکل ۴- روند کار GA-RF (الیان و گابر ۲۰۱۷).

معیارهای ارزیابی کارایی

$$SI = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (P_i - O_i)^2}}{\bar{O}}$$

$$0 \leq SI < \infty$$

[۱]

برای ارزیابی دقت روش‌های مورد مطالعه در برآورد تبخیر- تعرق مرجع، از پارامترهای ضریب همبستگی (CC)^۱، ضریب ویلموت (WI)^۲ و شاخص پراکندگی (SI)^۳ مطابق روابط ۱ تا ۳ استفاده شد.

$$WI = 1 - \left[\frac{\sum_{i=1}^n (O_i - P_i)^2}{\sum_{i=1}^n (|P_i - \bar{O}_i| + |O_i - \bar{O}_i|)^2} \right]$$

$$0 \leq WI \leq 1$$

[۲]

1. Correlation Coefficient
2. Willmott's Index of agreement
3. Scattered Index

نتایج و بحث

جهت بررسی و برآورد تبخیر- تعرق مرجع روزانه با استفاده از روش جنگل تصادفی در دو حالت منفرد و استفاده از الگوریتم ژنتیک برای بهینه‌سازی

متغیرهای مؤثر آن و انتخاب مناسب‌ترین ترکیب متغیرهای هواشناسی که دقت مناسبی در این امر دارند، ترکیب‌های متفاوت از داده‌های هواشناسی در شش سناریو مختلف متشکل از داده‌های متفاوت هواشناسی مطابق جدول ۴ تعریف گردید که سناریو ۶ برای نشان دادن دقت روش‌های جنگل تصادفی و جنگل تصادفی بهینه شده با الگوریتم ژنتیک در تخمین ET_0 است. در ادامه این سناریوها در سه ایستگاه استان آذربایجان شرقی (تبریز، سراب و مراغه) به‌کار گرفته شد.

$$CC = \frac{(\sum_{i=1}^n O_i P_i - \frac{1}{n} \sum_{i=1}^n O_i \sum_{i=1}^n P_i)}{(\sum_{i=1}^n O_i^2 - \frac{1}{n} (\sum_{i=1}^n O_i)^2)(\sum_{i=1}^n P_i^2 - \frac{1}{n} (\sum_{i=1}^n P_i)^2)} \quad [3]$$

$$-1 \leq CC \leq 1$$

در این روابط P_i تبخیر- تعرق گیاه مرجع برآورد شده با روش‌های مورد مطالعه، O_i تبخیر- تعرق گیاه مرجع محاسبه شده با روش فائو- پنمن- مانتیث، $\bar{O}_i$ میانگین تبخیر- تعرق گیاه مرجع محاسبه شده با روش فائو- پنمن- مانتیث و n تعداد داده‌ها می‌باشد. هرچه شاخص پراکندگی کمتر باشد نشان دهنده دقیق‌تر بودن روش می‌باشد. همچنین هرچه مقادیر ضریب ویلموت و ضریب همبستگی بیشتر و نزدیک به یک باشند، دقت روش بیشتر است.

جدول ۴- متغیرهای ورودی هر سناریو.

شماره سناریو	متغیرهای ورودی	نام مدل
۱	T_{min}, T_{max}	RF-1, GA-RF-1
۲	T_{min}, T_{max}, n	RF-2, GA-RF-2
۳	$T_{min}, T_{max}, RH_{min}, RH_{max}$	RF-3, GA-RF-3
۴	$T_{min}, T_{max}, RH_{min}, RH_{max}, U_2$	RF-4, GA-RF-4
۵	$T_{min}, T_{max}, RH_{min}, n, U_2$	RF-5, GA-RF-5
۶	$T_{min}, T_{max}, RH_{min}, RH_{max}, n, U_2$	RF-6, GA-RF-6

هستند، مورد ارزیابی قرار گرفت. با مبنا قرار دادن مدل FPM و پس از انجام محاسبات، دقت روش‌های یادگیری ماشین در مرحله صحت‌سنجی و روش‌های تجربی مورد مطالعه براساس معیارهای آماری ضریب همبستگی (CC)، شاخص پراکندگی (SI) و ضریب ویلموت (WI) محاسبه گردید (جدول ۵).

با نرم‌افزار Rapid Miner Studio به‌ترتیب ۷۵ و ۲۵ درصد از داده‌های دوره آماری مورد مطالعه برای واسنجی و صحت‌سنجی مدل‌های مذکور استفاده شد. همچنین کارایی روش‌های تجربی برآورد کننده ET_0 در سه دسته مبتنی بر دمای هوا، تابش و انتقال جرم که از نظر متغیرهای هواشناسی مطابق با سناریوهای ۱ تا ۵

جدول ۵- نتایج کلی مدل‌های RF، GA-RF در مرحله صحت‌سنجی و روش‌های تجربی در سه ایستگاه مورد مطالعه.

مدل	معیارهای آماری								
	تبریز			سراب			مراغه		
	WI	SI	CC	WI	SI	CC	WI	SI	CC
RF-1	۰/۹۷۶	۰/۲۱۷	۰/۹۵۴	۰/۹۷۶	۰/۲۶۲	۰/۹۲۵	۰/۹۷۷	۰/۲۰۸	۰/۹۵۷
GA-RF-1	۰/۹۷۶	۰/۲۱۶	۰/۹۵۶	۰/۹۷۶	۰/۲۵۸	۰/۹۲۹	۰/۹۷۹	۰/۱۹۸	۰/۹۶۱
RF-2	۰/۹۸۳	۰/۱۸۴	۰/۹۶۸	۰/۹۷۱	۰/۲۲۰	۰/۹۴۸	۰/۹۸۴	۰/۱۷۲	۰/۹۷۱
GA-RF-2	۰/۹۸۲	۰/۱۹۰	۰/۹۶۵	۰/۹۶۹	۰/۲۲۷	۰/۹۴۵	۰/۹۸۳	۰/۱۷۶	۰/۹۷۱
RF-3	۰/۹۸۰	۰/۱۹۸	۰/۹۶۳	۰/۹۶۹	۰/۲۲۶	۰/۹۴۶	۰/۹۸۱	۰/۱۹۰	۰/۹۶۴
GA-RF-3	۰/۹۷۹	۰/۲۰۴	۰/۹۶۰	۰/۹۶۷	۰/۲۳۳	۰/۹۴۲	۰/۹۸۰	۰/۱۹۳	۰/۹۶۳
RF-4	۰/۹۹۲	۰/۱۳۱	۰/۹۸۴	۰/۹۸۳	۰/۱۷۱	۰/۹۶۹	۰/۹۹۱	۰/۱۳۴	۰/۹۸۳
GA-RF-4	۰/۹۹۲	۰/۱۲۹	۰/۹۸۴	۰/۹۸۳	۰/۱۹۶	۰/۹۶۹	۰/۹۹۱	۰/۱۳۲	۰/۹۸۳
RF-5	۰/۹۹۴	۰/۱۱۰	۰/۹۸۹	۰/۹۸۸	۰/۱۴۶	۰/۹۷۸	۰/۹۹۴	۰/۱۰۷	۰/۹۸۹
GA-RF-5	۰/۹۹۵	۰/۱۰۷	۰/۹۸۹	۰/۹۸۹	۰/۱۴۱	۰/۹۷۹	۰/۹۹۴	۰/۱۰۵	۰/۹۸۹
RF-6	۰/۹۹۵	۰/۱۰۳	۰/۹۹۰	۰/۹۸۹	۰/۱۴۰	۰/۹۸۰	۰/۹۹۵	۰/۱۰۳	۰/۹۹۰
GA-RF-6	۰/۹۹۵	۰/۰۹۸	۰/۹۹۱	۰/۹۹۰	۰/۱۳۲	۰/۹۸۲	۰/۹۹۵	۰/۰۹۸	۰/۹۹۱
Hargreaves	۱/۰۷۵	۱/۰۷۵	۰/۹۵۴	۱/۷۶۲	۰/۵۸۸	۰/۹۴۹	۱/۷۵۸	۱/۰۳۲	۰/۹۵۵
Irmak-R _n	۰/۵۰۶	۰/۹۰۷	۰/۹۰۷	۰/۳۱۵	۰/۹۲۲	۰/۹۱۸	۰/۷۶۶	۰/۵۲۴	۰/۸۹۸
Romanenko	۰/۴۵۰	۰/۹۵۲	۰/۹۵۲	۰/۳۱۴	۰/۹۵۴	۰/۹۲۷	۰/۹۲۹	۰/۴۴۹	۰/۹۵۶
Meyer	۰/۹۴۰	۰/۹۶۴	۰/۹۶۴	۰/۴۹۳	۰/۹۱۶	۰/۹۴۷	۰/۸۱۱	۰/۹۶۵	۰/۹۶۹
Blaney-Cridell	۰/۲۲۰	۰/۹۸۴	۰/۹۸۴	۰/۳۱۶	۰/۹۶۲	۰/۹۷۸	۰/۹۸۵	۰/۱۹۴	۰/۹۸۷

پراکندگی RF-2 با پارامترهای هواشناسی دمای هوا و ساعات آفتابی، کمتر از شاخص پراکندگی RF-3 با پارامترهای دمای هوا و رطوبت نسبی است، بنابراین تابش خورشید نسبت به رطوبت نسبی از اهمیت بیشتری برخوردار است.

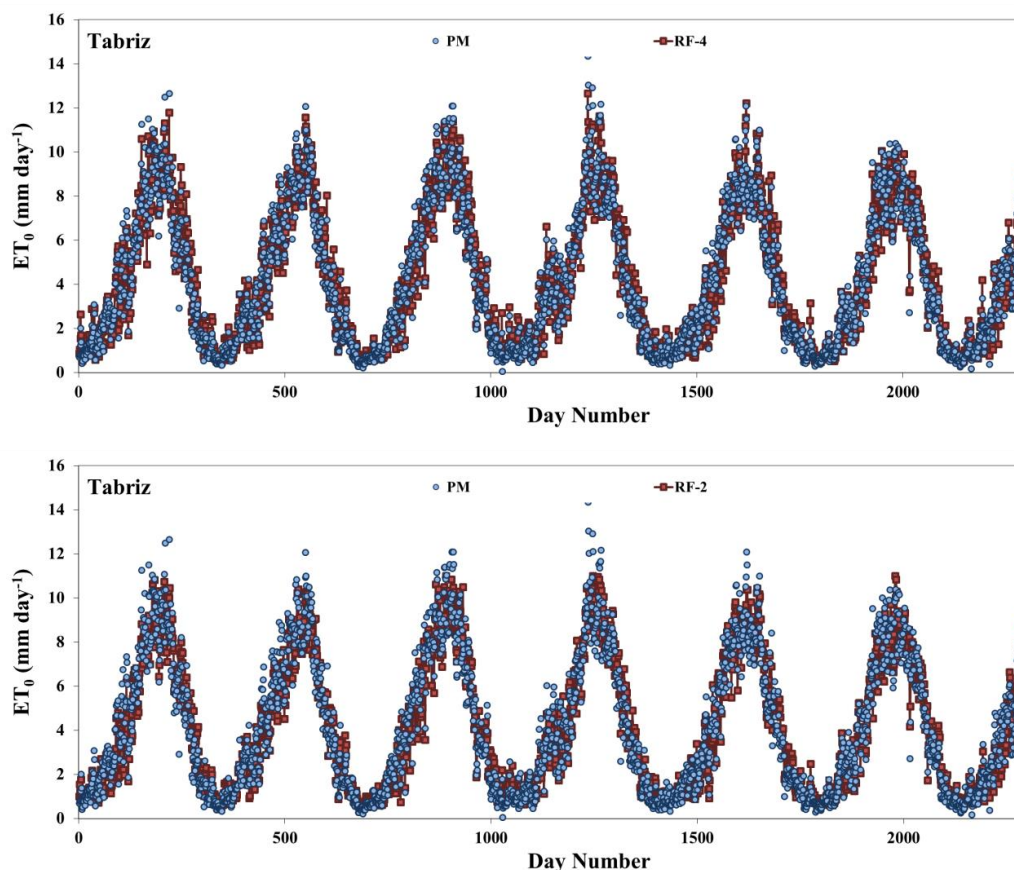
در این پژوهش متغیرهای ورودی هر سناریو برای یادگیری ماشین مطابق متغیرهای ورودی روش‌های تجربی است و در حالت مقایسه در هر سه ایستگاه جنگل تصادفی عملکرد بهتری نسبت به روش‌های تجربی نشان می‌دهد. روش بلانی-کریدل از روش‌های محاسباتی ET₀ بر پایه دمای هوا در ایستگاه تبریز با شاخص پراکندگی ۰/۲۲۰، در ایستگاه سراب با شاخص پراکندگی ۰/۳۱۶ و در ایستگاه مراغه با شاخص پراکندگی ۰/۱۹۴ بهترین عملکرد را در میان روش‌های تجربی دارد. نتایج روشن و همکاران (۲۰۱۱)

بر اساس نتایج جدول ۵، در ایستگاه‌های تبریز، سراب و مراغه RF-4 با ضریب همبستگی بین ۰/۹۶۹ تا ۰/۹۸۴، شاخص پراکندگی بین ۰/۱۳۱ تا ۰/۱۷۱ و ضریب وایلموت بین ۰/۹۸۳ تا ۰/۹۹۲ دقت بالایی دارد. در مرحله بعدی، سناریو ۲ تنها با متغیرهای ورودی دمای هوا و ساعات آفتابی با داشتن ضریب همبستگی ۰/۹۶۸، شاخص پراکندگی ۰/۱۸۴ و ضریب وایلموت ۰/۹۸۳ در ایستگاه تبریز، ضریب همبستگی ۰/۹۴۸، شاخص پراکندگی ۰/۲۲۰ و ضریب وایلموت ۰/۹۷۱ در ایستگاه سراب و ضریب همبستگی ۰/۹۷۱، شاخص پراکندگی ۰/۱۷۲ و ضریب وایلموت ۰/۹۸۴ در ایستگاه مراغه دقت قابل قبولی دارد. بنابراین به منظور تخمین تبخیر-تعرق مرجع روزانه با استفاده از متغیرهای ورودی کمتر سناریوی ۴ و ۲ در همه ایستگاه‌ها با دقت قابل قبولی می‌توانند مورد استفاده قرار گیرند. همچنین شاخص

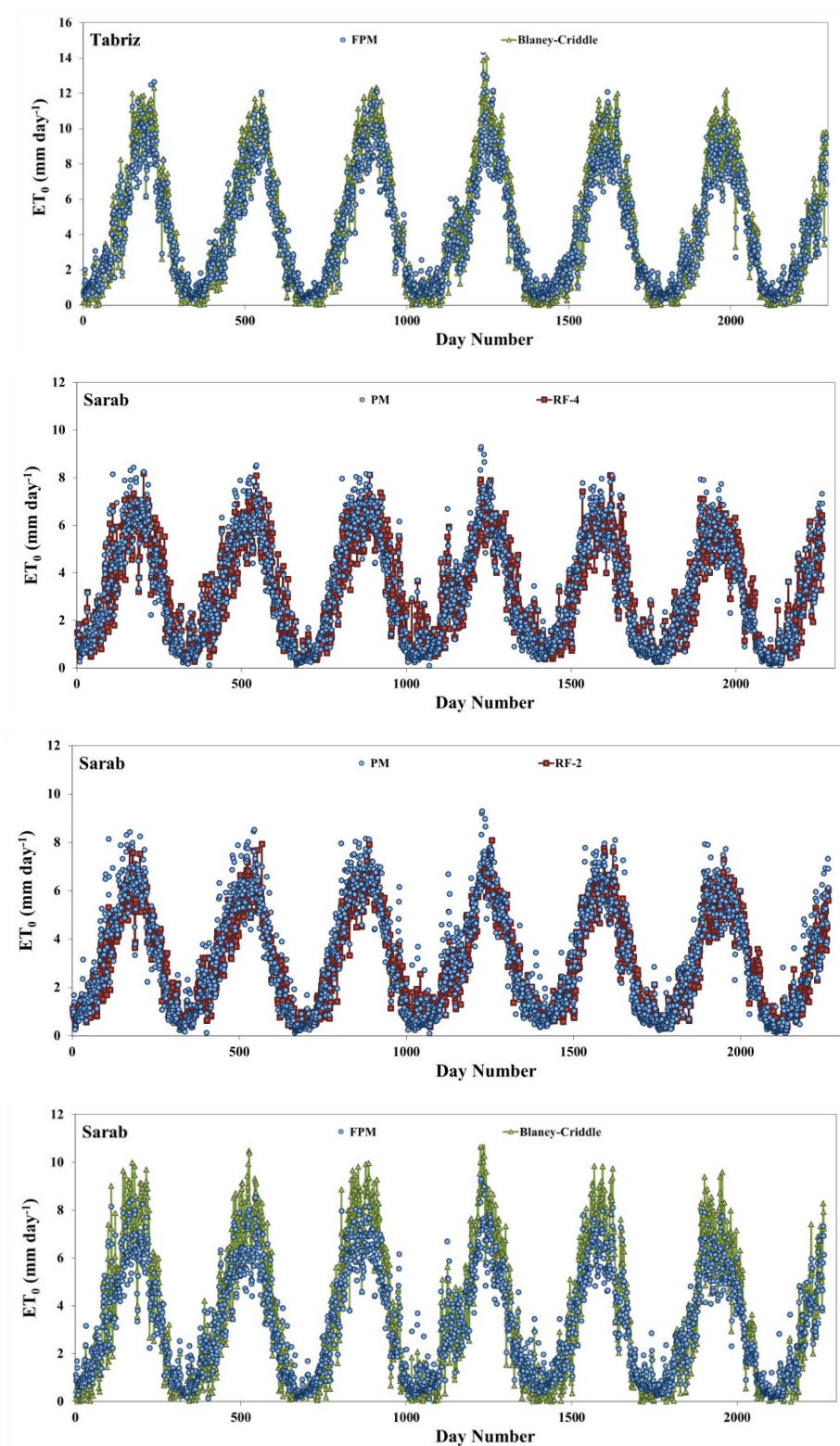
بنابراین می‌توان از روش جنگل تصادفی برای تخمین تبخیر- تعرق مرجع، با قبول اندکی خطا استفاده نمود.

در ادامه تغییرات زمانی مقادیر تبخیر- تعرق مرجع محاسبه شده با روش FPM و برآورد شده با روش‌های RF-2، RF-4 در مرحله صحت‌سنجی و روش دمایی بلانی- کریدل در شکل ۵ نشان داده شده است. همچنین نمودارهای پراکنش مقادیر مذکور در شکل ۶ قابل ملاحظه می‌باشد.

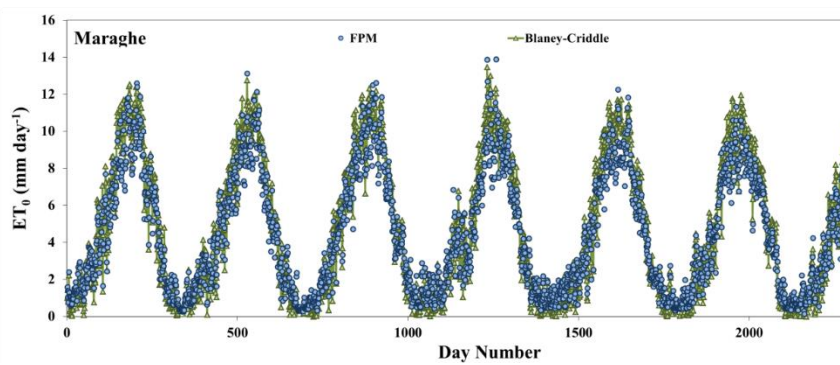
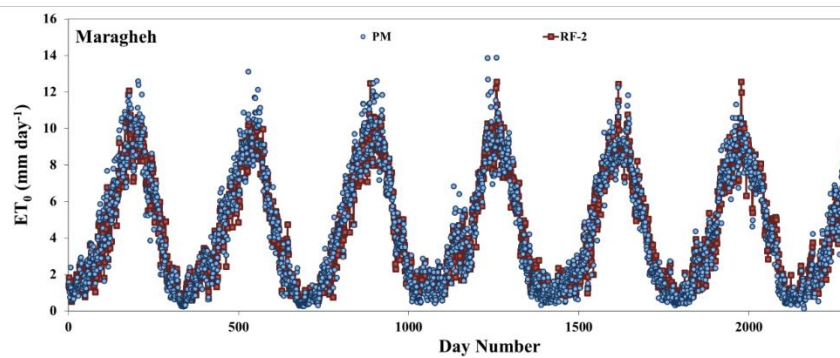
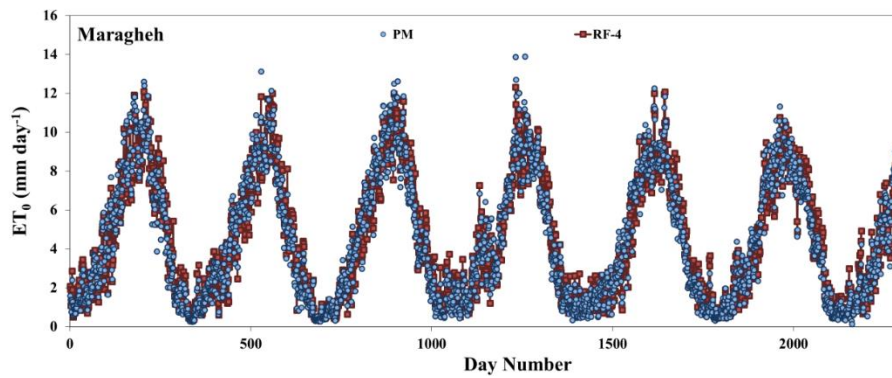
نیز همخوانی بهتر روش بلانی- کریدل نسبت سایر روش‌های تجربی با شرایط محیطی را بیان می‌نمود. روش بلانی- کریدل و سناریوی ۵ داده‌های ورودی مشابهی دارند اما روش یادگیری ماشین جنگل تصادفی دقیق‌تر است چراکه RF-5 در تبریز شاخص پراکندگی ۰/۱۱۰، در ایستگاه سراب شاخص پراکندگی ۰/۱۴۶ و در ایستگاه مراغه شاخص پراکندگی ۰/۱۰۷ دارد. از تحقیق چن و همکاران (۲۰۲۰) نیز نتایج مشابهی حاصل شده است. با توجه به نتایج RF و GA-RF، الگوریتم ژنتیک تأثیر قابل قبولی در بهبود عملکرد نداشت،



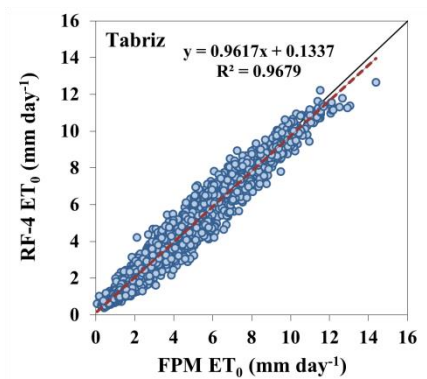
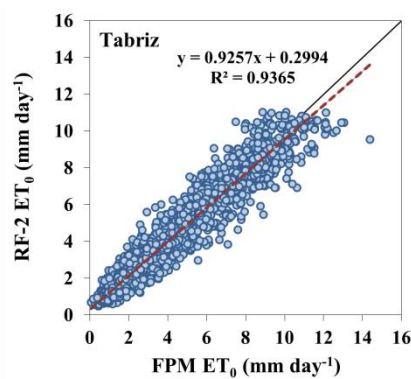
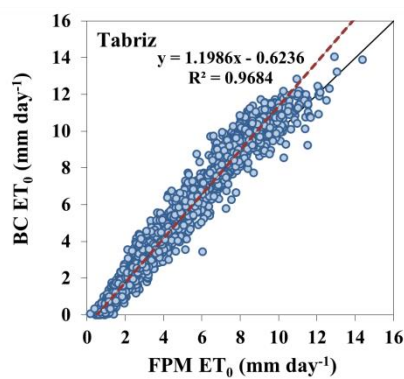
شکل ۵- تغییرات زمانی مقادیر ET_0 محاسبه شده با FPM و برآورد شده با روش‌های RF-2، RF-4 و بلانی- کریدل در سه ایستگاه مورد مطالعه.

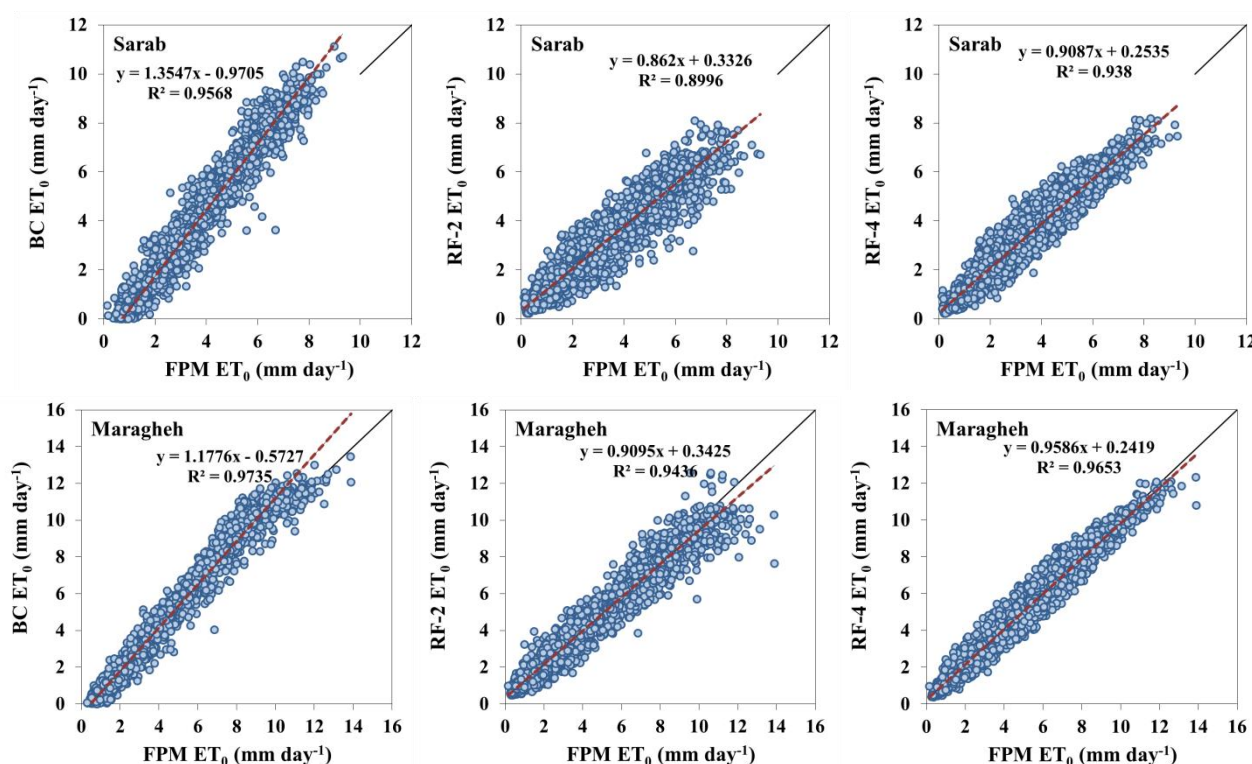


شکل ۵- ادامه



شکل ۵- ادامه





شکل ۶- نمودارهای پراکنش مقادیر ET_0 محاسبه شده با FPM و برآورد شده با RF-2، RF-4 و روش بلانی-کریدل در سه ایستگاه مورد مطالعه.

جنگل تصادفی^۷ (G) برای ایستگاههای تبریز، سراب و مراغه در جدول ۶ ارائه گردیده است.

از شکل ۶ استنباط می‌شود که در هر سه ایستگاه RF-4 دقت بالایی در برآورد ET_0 دارد. RF-2 کم‌برآوردی و روش بلانی-کریدل بیش‌برآوردی دارد. همچنین متغیرهای پیش‌فرض روش جنگل تصادفی و جنگل تصادفی بهینه‌سازی شده با الگوریتم ژنتیک شامل تعداد درخت‌های جنگل تصادفی^۱ (A)، عمق حداکثر جنگل تصادفی^۲ (B)، اطمینان جنگل تصادفی^۳ (C)، اندازه برگ حداقل جنگل تصادفی^۴ (D)، حداقل اندازه برای تقسیم جنگل تصادفی^۵ (E)، تعداد گزینه‌های قبل از هرس جنگل تصادفی^۶ (F) و نسبت زیر مجموعه

1. Random Forest number of trees
2. Random Forest maximal depth
3. Random Forest confidence
4. Random Forest minimal leaf size
5. Random Forest minimal size for split
6. Random Forest number of pre pruning alternatives

جدول ۶- متغیرهای مدل RF و GA-RF برای ایستگاه‌های مورد مطالعه.

ایستگاه	مدل	پارامتر					
		A	B	C	D	E	F
تبریز	RF	۱۰۰	۱۰	۰/۱۰۰	۲	۴	۳
	GA-RF-1	۹۴	۵	۰/۳۳۶	۴۱	۲۰	۳
	GA-RF-2	۸۱	۸۶	۰/۳۵۵	۴۱	۴۹	۱۸
	GA-RF-3	۸۱	۳۶	۰/۲۲۲	۴۱	۲۰	۳
	GA-RF-4	۷۸	۶۵	۰/۱۴۸	۴۱	۲۰	۳
	GA-RF-5	۸۱	۳۶	۰/۲۲۹	۱	۸۰	۵۵
	GA-RF-6	۸۱	۳۶	۰/۱۹۱	۱	۸۰	۵۵
سراب	RF	۱۰۰	۱۰	۰/۱۰۰	۲	۴	۳
	GA-RF-1	۹۴	۵	۰/۳۴۸	۴۱	۲۰	۳
	GA-RF-2	۸۱	۳۶	۰/۲۳۳	۱	۸۰	۵۵
	GA-RF-3	۷۸	۶۵	۰/۱۴۸	۴۱	۲۰	۳
	GA-RF-4	۹۴	۳۶	۰/۲۱۳	۱	۸۰	۳
	GA-RF-5	۹۴	۳۶	۰/۲۰۲	۱	۸۰	۳
	GA-RF-6	۹۴	۳۶	۰/۲۹۲	۱	۲۰	۳
مراغه	RF	۱۰۰	۱۰	۰/۱۰۰	۲	۴	۳
	GA-RF-1	۸۱	۵	۰/۳۴۹	۴۱	۲۰	۵۵
	GA-RF-2	۸۱	۵	۰/۳۴۵	۴۱	۲۰	۵۵
	GA-RF-3	۹۴	۵	۰/۳۶۵	۴۱	۲۰	۳
	GA-RF-4	۹۴	۳۶	۰/۲۱۱	۱	۸۰	۳
	GA-RF-5	۹۴	۳۶	۰/۲۳۹	۱	۸۰	۳
	GA-RF-6	۹۴	۳۶	۰/۲۳۱	۱	۸۰	۳

نتیجه‌گیری کلی

بخش کشاورزی بزرگترین مصرف کننده آب می‌باشد و با توجه به کاهش منابع آب در دسترس، برآورد دقیق میزان آب مورد نیاز این بخش کمک شایانی به مدیریت منابع آب خواهد کرد. در این مطالعه به برآورد مقادیر تبخیر- تعرق مرجع به‌عنوان یکی از متغیرهای اقلیمی مهم در سه ایستگاه تبریز، سراب و مراغه در بازه زمانی ۱۳۸۰ تا ۱۴۰۰ و با تعریف شش

سناریو متشکل از ترکیب‌های متفاوتی از متغیرهای هواشناسی، پرداخته شد. کاربرد روش جنگل تصادفی در دو حالت منفرد و استفاده از الگوریتم ژنتیک برای بهینه‌سازی متغیرهای مؤثر آن و نیز سه گروه روش‌های تجربی برآورد کننده تبخیر- تعرق مرجع، شامل روش‌های بر پایه دمای هوا (هارگریوز- سامانی، بلانی- کرایدل و روماننکو)، تابش (ایرماک) و انتقال جرم (مییر) با در نظر گرفتن مدل فائو- پنمن- مانتیث

به‌عنوان روش مبنا، مورد بررسی قرار گرفتند. در نهایت، با استفاده از معیارهای آماری به ارزیابی عملکرد هر یک از روش‌های مذکور پرداخته شد. بر اساس نتایج حاصل، به‌منظور تخمین ET_0 با استفاده از متغیرهای هواشناسی کمتر سناریو ۴ و ۲ در همه ایستگاه‌ها با دقت قابل قبولی می‌توانند مورد استفاده قرار گیرند. در حالت مقایسه در هر سه ایستگاه جنگل

تصادفی عملکرد بهتری نسبت به روش‌های تجربی نشان داد. همچنین از روش‌های تجربی مورد مطالعه، بلانی- کریدل بیشترین دقت و هارگریوز- سامانی کمترین دقت را داشت. در نهایت استفاده از روش RF به‌منظور برآورد دقیقی از تبخیر- تعرق مرجع در استان آذربایجان شرقی پیشنهاد گردید.

منابع مورد استفاده

- Allen RG, Pereira LS, Raes D and Smith M, 1998. Crop evapotranspiration-Guidelines for computing crop water requirements. FAO Irrigation and Drainage paper 56.
- Bayat VM, Zare AH, Maroufi S, Sabziparvar A and Soltani F, 2009. Simulation of reference evapotranspiration using artificial neural method and empirical methods and comparison with experimental lysimeter data in cold semi-arid climate of Hamedan. Water and Soil Conservation Research (Agricultural Sciences and Natural Resources) 16(4): 79-100. (In Persian with English abstract)
- Blaney HF and Criddle WD, 1950. Determining requirements water in irrigated areas from climatological and irrigation data. Washington Soil Conservation Service 48.
- Chen H, Huang JJ and McBean E, 2020. Partitioning of daily evapotranspiration using a modified shuttleworth-wallace model, random forest and support vector regression, for a cabbage farmland. Agricultural Water Management 228: 105923.
- Da Silva Junior JC, Medeiros V, Garrozi C, Montenegro A and Gonçalves GE, 2019. Random forest techniques for spatial interpolation of evapotranspiration data from Brazilian's Northeast. Computers and Electronics in Agriculture 166: 105017.
- Elyan E and Gaber MM, 2017. A genetic algorithm approach to optimising random forests applied to class engineered data. Information Sciences 384: 220-234.
- Emadi A, Mohammadiha A and Mohammad Vali Samani J, 2011. Mathematical model for auto-calibration of area-reduction method in sediment distribution in of dam reservoir using genetic algorithm. Journal of Water and Soil Science (Agricultural Sciences and Industries) 25(2): 356-364. (In Persian with English abstract)
- Fan J, Yue W, Wu L, Zhang F, Cai H, Wang X and Xiang Y, 2018. Evaluation of SVM, ELM and four tree-based ensemble models for predicting daily reference evapotranspiration using limited meteorological data in different climates of China. Agricultural and Forest Meteorology 263: 225-241.
- Feng Y, Cui N, Gong D, Zhang Q and Zhao L, 2017. Evaluation of random forests and generalized regression neural networks for daily reference evapotranspiration modelling. Agricultural Water Management 193: 163-173.
- Hargreaves GH and Samani ZA, 1985. Reference crop evapotranspiration from temperature. Applied Engineering in Agriculture 1(2): 96-99.
- Hashemian AH, Manochchri S, Afshari D, Manochchri Z, Salari N and Shahsavari S, 2019. Prognosis of multiple sclerosis disease using data mining approaches random forest and support vector machine based on genetic algorithm. Tehran University Medical Journal 77(1): 33-40. (In Persian with English abstract)
- Irmak S, Irmak A, Allen RG and Jones JW, 2003. Solar and net radiation-based equations to estimate reference evapotranspiration in humid climates. Journal of Irrigation and Drainage Engineering 129(5): 336-347.
- Karimi S, Shiri J and Nazemi AH, 2013. Estimating daily reference crop evapotranspiration using artificial intelligences-based ANFIS and ANN techniques and empirical models. Water and Soil Science 23(2): 139-158. (In Persian with English abstract)

- Kouchami Sardoo I, Shirani H and Besalatpour AA, 2020. Determining the features influencing the structural stability of soils of arid regions using a hybrid GA-ANN algorithm. *Applied Soil Research* 8(3): 129-143. (In Persian with English abstract)
- Landeras G, Ortiz-Barredo A and López JJ, 2008. Comparison of artificial neural network models and empirical and semi-empirical equations for daily reference evapotranspiration estimation in the Basque Country (Northern Spain). *Agricultural Water Management* 95(5): 553-565.
- Meyer A, 1926. U ber einige Zusammenhänge zwischen Klima und Boden in Europa. *Chemie der Erde* 2: 209-347.
- Mohammadi M, Mousavi F, Farzin S and Karami H, 2019. Optimal operation of Salman Farsi dam reservoir using Wall algorithm and its hybrid with genetic algorithm based on multi-criteria decision method. *Echo Hydrology* 6 (2): 281-293. (In Persian with English abstract)
- Norouzi H and Nadiri A, 2018. Groundwater level prediction of Boukan Plain using fuzzy logic, random forest and neural network models. *Journal of Range and Watershed Managment* 71(3): 829-845. (In Persian with English abstract)
- Oghnoum M, Fegghi J, Makhdoum M, Moghaddamnia A and Etemad V, 2019. Land capability evaluation of afforestation using random forest algorithm (Kan Watershed, Tehran). *Journal of Forest Research and Development* 5(3): 387-403. (In Persian with English abstract)
- Onnabi Milani A and Neyshabouri MR, 2018. Comparison of some empirical estimating methods of reference evapotranspiration in Tabriz Plain using lysimeter and proposing a model for its determination from climatic data. *Water and Soil Science* 28(1): 41-54. (In Persian with English abstract)
- Piri H and Poozan M, 2019. Evaluation of 24 models of reference evapotranspiration in different climates of Iran. *Iranian Journal of Ecohydrology* 6(3): 611-622. (In Persian with English abstract)
- Romanenko VA, 1961. Computation of the autumn soil moisture using a universal relationship for a large area. *Ukrainian Hydrometeorological Research Institute* 3: 12-25.
- Roshan GR, Khoshakhlagh F and Karampur M, 2012. Assessing, modifying and synthesizing a suitable model for estimation of potential evapotranspiration in Iran. *Physical Geography Research Quarterly* 43(78): 49-68. (In Persian with English abstract)
- Salam R and Islam ARMT, 2020. Potential of RT, Bagging and RS ensemble learning algorithms for reference evapotranspiration prediction using climatic data-limited humid region in Bangladesh. *Journal of Hydrology* 590: 125241.
- Samadianfard S, Salarifar M, Javidan S and Mikaeili F, 2020. Estimation of daily reference evapotranspiration in humid climates using data-driven methods of gaussian process regression, support vector regression and random forest. *Environment and Water Engineering* 6(4): 360-373. (In Persian with English abstract)
- Shiri J, Zounemat-Kermani M, Kisi O and Mohsenzadeh Karimi S, 2020. Comprehensive assessment of 12 soft computing approaches for modelling reference evapotranspiration in humid locations. *Meteorological Applications* 27(1):1-15.
- Siasar H, Honar T and Abdolahi-pour M, 2020. Comparing of generalized linear models, random forest and gradient boosting trees in estimation of reference crop evapotranspiration (Case study: The Sistan Plain). *Journal of Water and Soil Science (Journal of Science and Technology of Agriculture and Natural Resources)* 23(4): 395-410. (In Persian with English abstract)
- Wang S, Lian J, Peng Y, Hu B and Chen H, 2019. Generalized reference evapotranspiration models with limited climatic data based on random forest and gene expression programming in Guangxi, China. *Agricultural Water Management* 221: 220-230.
- Zare Abyaneh H, 2014. Modeling and forecasting of reference crop evapotranspiration using time series (Case study: Kermanshah Province). *Journal of Agricultural Meteorology* 2(1): 45-56. (In Persian with English abstract)